9К

УДК 681.3

М.В. Крыжановский, М.Ю. Мальсагов
ЦОНТ НИИ системных исследований РАН, г. Москва, Российская Федерация
iont.niisi@gmail.com

Обобщение процедуры клиппирования в задачах оптимизации в дискретном пространстве*

Исследована возможность применения процедуры клиппирования в задаче оптимизации квадратичного функционала $E = (\mathbf{x}, \mathbf{Ax})$. Показано, что непосредственное применение процедуры клиппирования не дает особого выигрыша в ускорении работы алгоритма при поиске глобального минимума. Предложена модификация процедуры клиппирования с параметром q (число градаций). Показано, что с увеличением q вероятность совпадения направления градиентов $E(x)$ и его клиппированного аналога $E_C(x) = (\mathbf{x}, \mathbf{Cx})$ возрастает до 1.

Введение

Впервые замена матрицы **A** на клиппированную матрицу **C** исследовалась в задачах распознавания образов [1], [2]. Были получены аналитические оценки емкости нейросетевой памяти и ее распознающей способности. Эти исследования были продолжены в [3-5], основные результаты которых приведены ниже:

– уменьшение энергии $E_C(x)$ клиппированной сети Хопфилда при переходе из одного состояния в другое сопровождается уменьшением энергии $E(x)$ исходной сети;

– быстродействие алгоритма, основанного на использовании клиппированной матрицы, в 40 раз превышает быстродействие алгоритма, основанного на использовании нейронной сети Хопфилда;

– приблизительно во столько же раз уменьшаются требования к оперативной памяти.

На основе этого в [6], [7] было предложено использование процедуры клиппирования при решении задач оптимизации. **В работе предложен** модифицированный алгоритм клиппирования, позволяющий ускорить поиск глобального минимума.

Применение традиционной процедуры клиппирования

Поиск глобального минимума квадратичного функционала $E = (\mathbf{x}, \mathbf{Ax})$ в дискретном бинарном пространстве заключается в многократном применении стандартной Хопфилдовой процедуры оптимизации, т.е. проведении серии «спусков» по энергетической поверхности $E(x)$ из начальных состояний $\{x_1\}$ в конечные $\{x_4\}$, переходы $\{x_1\} \to \{x_4\}$. Отбор наиболее глубокого минимума проводится в ходе серии спусков.

При использовании клиппированного функционала процесс поиска разбивается на 2 этапа, переходы $\{x_1\} \to \{x_2\} \to \{x_3\}$. Найденные на 1-м этапе, локальные минимумы $\{x_2\}$ функционала $E_C(x)$ становятся исходными стартовыми при минимизации функционала $E(x)$ на 2-м.

* Работа выполнена при поддержке гранта РФФИ 09-07-00159-а

Для определения эффективности такого подхода было проведено компьютерное моделирование, в ходе которого генерировалась матрица $\mathbf{A}$ размерности 100 со случайным равномерным распределением элементов и вычислялся ее клиппированный аналог – матрица $\mathbf{C}$. Для каждой конфигурации нейронной сети проводилось 200 000 стартов из состояний, задаваемых вектором $\mathbf{x}$, компоненты которого генерировались случайным образом. В ходе эксперимента для каждого «спуска» определялся объем вычислений и фиксировалась энергия локального минимума.

На рис. 1 представлены результаты сравнения двух методов оптимизации – стандартного и с применением клиппирования. По оси абсцисс отложена «энергия» $\varepsilon = (E_0 - E)/E_0$, где E_0 – энергия глобального минимума, E – энергия полученного локального минимума функционала $E = (\mathbf{x}, \mathbf{Ax})$. По оси ординат отложена плотность вероятности нахождения минимума.

Кривая 1 – отображает распределение по «энергиям» ε исходных стартовых состояний $\{x_1\}$.

Кривая 2 – характеризует распределение состояний $\{x_2\}$, полученных при оптимизации клиппированного функционала $E_C(x)$, т.е. переходы $\{x_1\} \to \{x_2\}$.

Кривая 3 – определяет распределение $\{x_2\} \to \{x_3\}$ после коррекции состояний сетью Хопфилда на втором этапе.

Кривая 4 – соответствует распределению по «энергии» ε при переходах $\{x_1\} \to \{x_4\}$ используя сеть Хопфилда.

Из рис. 1 видно, что использование клиппированной сети действительно смещает распределение состояний по «энергиям» ε на 0,7 по сравнению с исходным $\{x_1\}$. Несмотря на такой сдвиг, объем вычислений уменьшился незначительно. Конечное распределение с использованием двухэтапного алгоритма близко к распределению с использованием стандартной нейронной сети Хопфилда. Одинаковы и вероятности попадания в глобальный минимум, равные $\approx 0{,}006$. Для увеличения скорости работы 2-этапного алгоритма поиска более глубоких минимумов и был предложен модифицированный алгоритм клиппирования.

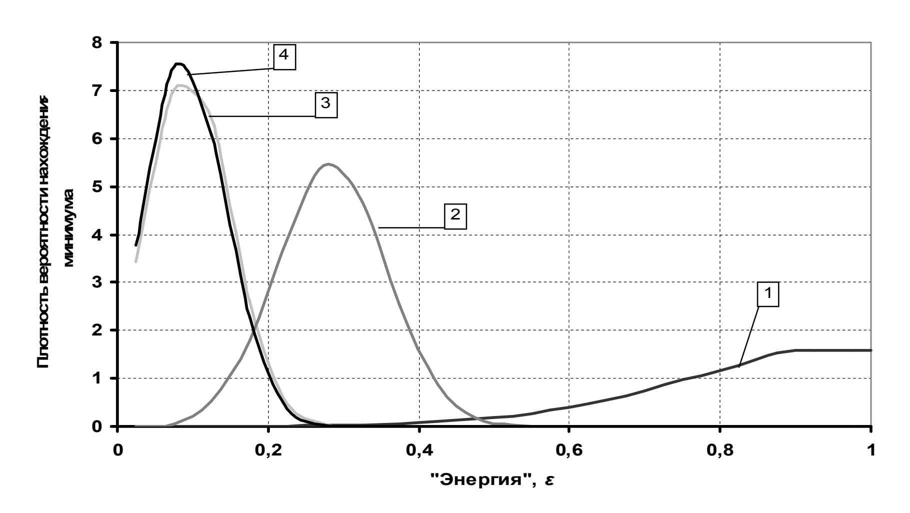


Рисунок 1 – Сравнение двух методов оптимизации: на основе сети Хопфилда и метода, использующего клиппированную сеть

Применение модифицированной процедуры клиппирования при поиске глобального минимума

Модификация процедуры клиппирования заключается в следующем. Каждому элементу матрицы связей **A** нейронной сети Хопфилда сопоставляется элемент матрицы **C** по формуле:

$$C_{ik} = \frac{1}{(q+1/2)}\text{sgn}(A_{ik})\cdot round\left[(q+1/2)\cdot|A_{ik}|\right], \quad (1)$$

где q – число градаций, больше нуля. Функция sgn дает знак числа, а функция $round$ производит округление до ближайшего целого.

Будем анализировать корреляцию градиентов $\mathbf{H_A}$ и $\mathbf{H_C}$, т.е. исходного функционала $E(x)$ и клиппированного $E_C(x)$, которые имеют вид:

$$\mathbf{H_A} = \mathbf{A}\cdot\mathbf{x}, \quad (2)$$

$$\mathbf{H_C} = \mathbf{C}\cdot\mathbf{x} = \mathbf{A}\cdot\mathbf{x} + \mu\cdot\mathbf{B}\cdot\mathbf{x}. \quad (3)$$

Величина $\mu = 1/(2q+1)$ характеризует степень огрубления. Второе слагаемое в (3) характеризует остаток, получаемый от огрубления элементов **A**. **B** – матрица с равномерным случайным распределением элементов в диапазоне $[-1;1]$, а каждая компонента вектора $(\mathbf{B},\mathbf{x})$ ведет себя как случайная величина, имеющая Гауссово распределение $(N >> 1)$. Вычислим вероятность совпадения направления полей P, т.е. совпадение по знаку каких-то компонент векторов $\mathbf{H_A}$ и $\mathbf{H_C}$:

$$P = \Pr[H_A \cdot H_C > 0]. \quad (4)$$

Элементы исходной матрицы **A** равномерно распределены с нулевым средним $\bar{a}$ и дисперсией $\sigma^2(a)$.

С учетом выражений (2) и (3) можно показать, что математическое ожидание и дисперсия величин H_A и H_C описываются выражениями:

$$\bar{H}_A = 0, \quad \sigma^2(H_A) = n\sigma^2(a); \quad (5)$$

$$\bar{H}_C = 0, \quad \sigma^2(H_C) = (1-\mu^2)\cdot\sigma^2(H_A); \quad (6)$$

$$\overline{H_C \cdot H_A} = \sigma^2(H_C). \quad (7)$$

С учетом выражений (5) – (7) коэффициент корреляции ρ градиентов H_A и H_C будет равен:

$$\rho = \frac{\overline{H_A \cdot H_C} - \bar{H}_A\cdot\bar{H}_C}{\sigma(H_A)\sigma(H_C)} \cong 1 - \frac{1}{2}\mu^2. \quad (8)$$

Минимально достигаемое значение $\rho = 0{,}944$ соответствует $q = 1$ и $\mu = 1/3$.

В свою очередь, вероятность совпадения направления полей P определяется как:

$$\Pr(H_A \cdot H_C > 0) = \frac{1}{\pi\sigma(H_A)\sigma(H_C)\sqrt{1-\rho^2}}\int_0^\infty\int_0^\infty Exp\left\{-\frac{1}{2(1-\rho^2)}\left[f(H_A,H_C)\right]\right\}dH_A dH_C, \quad (9)$$

где $f(H_A,H_C) = \left(\frac{H_A-\bar{H}_A}{\sigma(H_A)}\right)^2 - 2\rho\left(\frac{H_A-\bar{H}_A}{\sigma(H_A)}\right)\left(\frac{H_C-\bar{H}_C}{\sigma(H_C)}\right) + \left(\frac{H_C-\bar{H}_C}{\sigma(H_C)}\right)^2.$

Вычисление этого двойного интеграла возможно только численно. Результаты расчета приведены на рис. 2 .

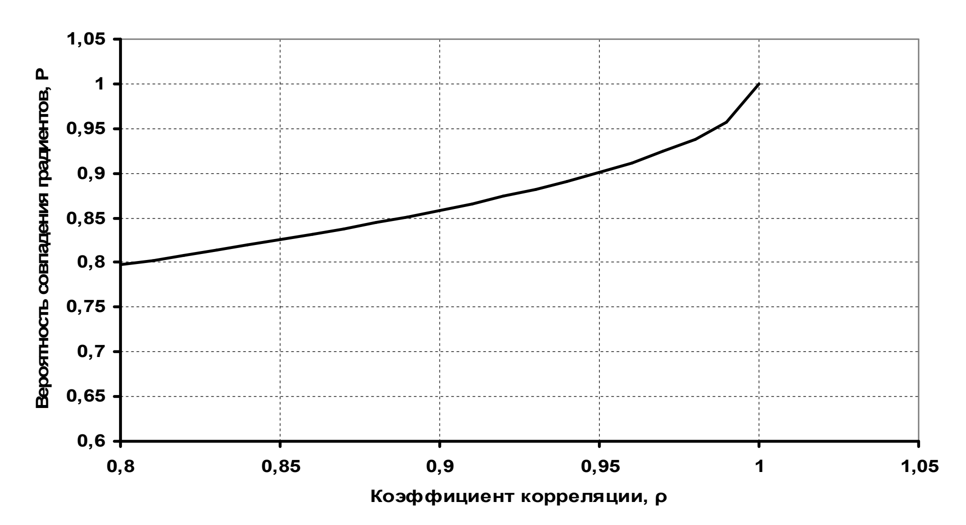
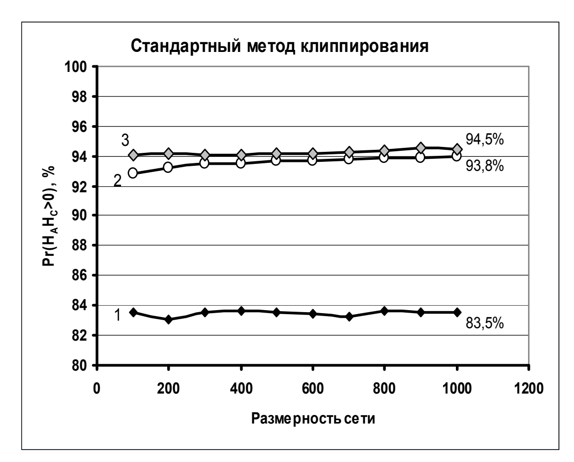


Рисунок 2 – Вероятность совпадения градиентов клиппированного и исходного функционалов в зависимости от коэффициента корреляции

На основании вычислений формулы (9) вероятность совпадения градиентов можно приближенно оценить по формуле (10), справедливой при значениях коэффициента корреляции ρ, близких к 1:

$$P(\mu) = 1 - \frac{3}{2}\mu^2 . \qquad (10)$$

Из приведенного на рис. 2 графика, формул (9) и (10) следует, что локальные градиенты исходного и клиппированного функционалов с большой вероятностью совпадают $(P \geq 0,894)$. С ростом числа градаций q (уменьшение μ) эта вероятность возрастает и стремится к 1. Поскольку процесс оптимизации заключается в последовательном перевороте всех N спинов модели Хопфилда, то становится очевидным, что с ростом размерности задачи $(N >> 1)$ асимптотически стремится к нулю вероятность того, что в процессе оптимизации функционала энергия E повысится.



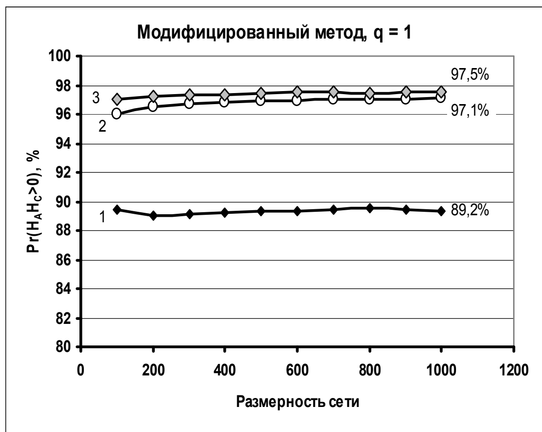


Рисунок 3 – Вероятность совпадения градиентов при использовании обычной (слева) и модифицированной (справа) процедуры клиппирования

Для проверки полученных результатов было проведено компьютерное моделирование. На рис. 3 представлены экспериментальные данные о вероятности совпадений градиентов $\mathbf{H}_A$ и $\mathbf{H}_C$ в случае применения обычной процедуры клиппирования (слева) и его модификации при $q = 1$ (справа). Для этого случайным образом выбиралась точка x_1 (кривая 1), из которой сеть Хопфилда с клиппированной матрицей межсвязей $\mathbf{C}$ конвергировала в ближайший локальный минимум x_2 (кривая 2). Точка минимума x_3 получена после коррекции состояний стандартной сетью Хопфилда (кривая 3). В этих точках x_1, x_2, x_3 определялось соответствие направлений векторов $\mathbf{H}_A$ и $\mathbf{H}_C$. По оси ординат отложена размерность сети, для которой проводились измерения.

Полученные результаты экспериментов точно согласуются с теорией. Так, измеренная экспериментально вероятность совпадения градиентов для модифицированного метода при $q = 1$ дает $P = 0,892$ (справа, кривая 1), в то же время расчетное значение составляет $P = 0,896$.

Сопоставление экспериментальных данных на этих же рисунках показывает преимущество модифицированного метода клиппирования.

1. В случайно выбранных точках старта вероятность совпадения для модифицированного метода составляет $P = 0,892$, в то время как для исходного – $P = 0,835$.

2. В точках минимума клиппированной сети вероятность слева составляет $P = 0,938$, справа – $P = 0,971$.

3. Соответственно в точках минимума стандартной сети слева получаем $P = 0,945$, а справа – $P = 0,975$.

Полученные результаты не зависят от размерности сети.

На рис. 4 показано соотношение между «энергиями» минимумов клиппированной сети и стандартной сети Хопфилда для различных значений параметра q. Для этого выбиралась случайным образом точка x_0, из которой сеть Хопфилда с матрицей межсвязей $\mathbf{C}(q)$ конвергировала в ближайший локальный минимум x_m. Рассчитывались значения $E(x_m)$ и клиппированного $E_C(x_m)$ функционалов и далее приведенные значения «энергий» ε.

Видно, что с увеличением параметра q область пропорциональной зависимости становится все больше, ее граница становится ближе к началу координат. Это означает, что с ростом параметра q область соответствия «глубже клиппированный минимум – глубже минимум стандартной сети» все больше приближается к глобальному минимуму.

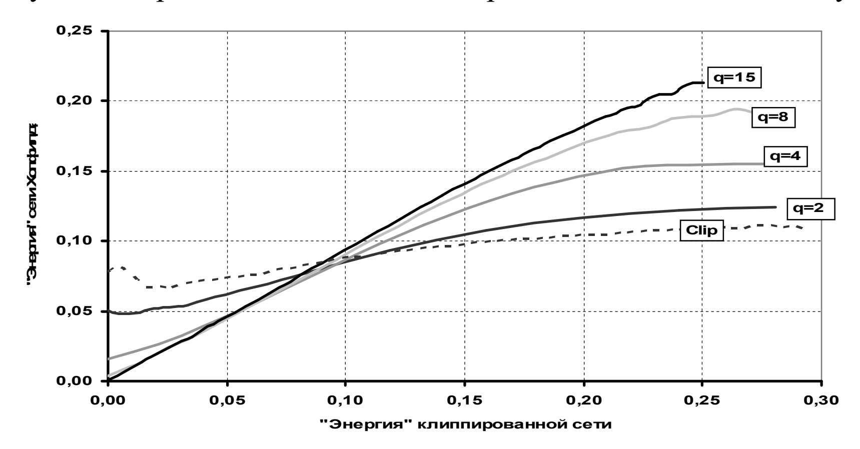


Рисунок 4 – Соотношение между «энергиями» минимумов клиппированной сети и стандартной сетью Хопфилда для различных значений параметра q

На рис. 5 представлены распределения состояний по энергии для разного числа градаций, полученные после первого этапа оптимизации (процедура q-клиппирования). Видно, что с увеличением величины q распределение смещается влево, тем самым приближая нас к глобальному минимуму и уменьшая путь, который необходимо проделать стандартной сети Хопфилда на 2-м этапе. Стоит отметить, что модификация процедуры клиппирования не изменила вероятности нахождения глобального минимума.

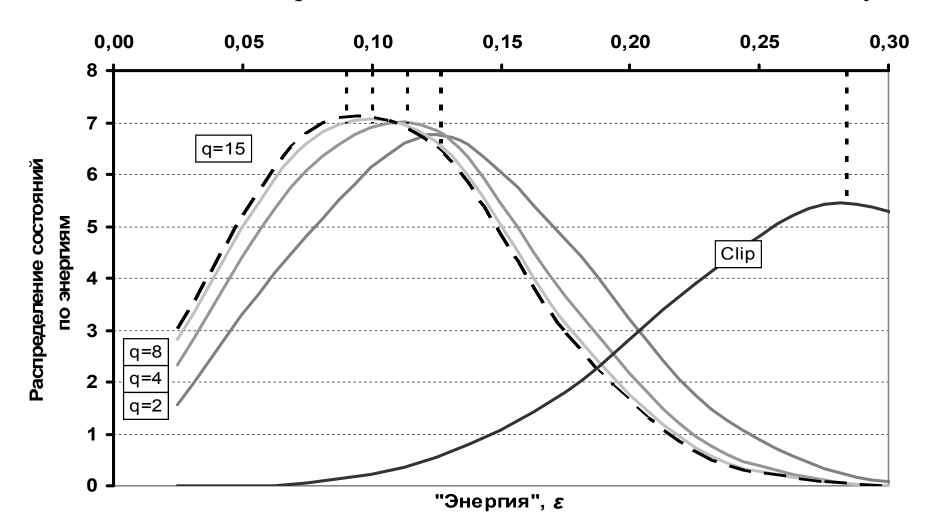


Рисунок 5 – Распределение состояний по энергии для разного числа градаций

Применение модифицированного алгоритма клиппирования

Основной объем вычислений при расчетах нейронной сети Хопфилда приходится на вычисление градиентов, т.е. на операции умножения матрицы на вектор. Для матриц высокой размерности $N \sim 10^3 - 10^4$ используется 10-байтное представление чисел для получения необходимой точности вычислений. Представление матрицы межсвязей нейронной сети с помощью чисел укороченной разрядности позволяет ускорить вычислительный процесс. Так, если сложение 2-х 10-байтных чисел требует 1-го такта процессорного времени, то для сложения 2-х десятимерных векторов, компоненты которых 1-байтные числа, потребуется времени меньше 1 такта. Кроме того, загрузка операндов из памяти в регистры процессора также требует времени, сопоставимого со временем выполнения операции. Поэтому «эффективное» время выполнения 1-байтовой операции в ~15 раз меньше, чем 10-байтовой.

Идея модификации заключается в применении целых чисел в одно- и двухбайтовом представлении для матричных элементов. Соответственно этому и был построен алгоритм спуска по «энергетической поверхности», который состоит из 3-х этапов, когда локальный минимум, полученный на текущем этапе, являлся начальным приближением для следующего этапа. На первом и втором этапах использовалась модификация клиппирования с числом градаций $q = 2 \div 64$ на 1-м (однобайтные операции) и $q \geq 255$ на 2-м (двухбайтные операции). Для коррекции состояний на третьем этапе использовалась стандартная сеть Хопфилда.

Ускорение вычислительного процесса зависит от числа градаций $q^{(1)}$, выбранного на первом этапе. На втором этапе число q может быть выбрано максимально большим $q \geq 2048$.

Для определения оптимального значения $q^{(1)}$ на 1-м этапе и соответственно величины ускорения алгоритма θ был проделан вычислительный эксперимент. На каждом этапе оптимизации функционала определялось число итераций при попадании в локальный минимум, которое пропорционально объему вычислений. При этом фиксировалось количество шагов при «спуске» с использованием исходной сети Хопфилда.

Результаты вычислительного эксперимента приведены на рис. 6 и в табл. 1. Ускорение θ работы алгоритма определяется формулой (11):

$$\theta = \frac{15 I^{(H)}}{I^{(1)} + 2I^{(2)} + 10 I^{(HL)}}, \quad (11)$$

где $I^{(H)}$ – количество итераций, используя стандартный метод Хопфилда;

$I^{(1)}$ – количество итераций на первом этапе (однобайтная арифметика);

$I^{(2)}$ – количество итераций на втором этапе (двухбайтные вычисления);

$I^{(HL)}$ – количество итераций на этапе коррекции алгоритма стандартным методом Хопфилда.

На рис. 6 приведено соотношение между числом итераций, затрачиваемых на разных этапах модифицированного алгоритма клиппирования по отношению к числу итераций, затрачиваемых стандартным методом.

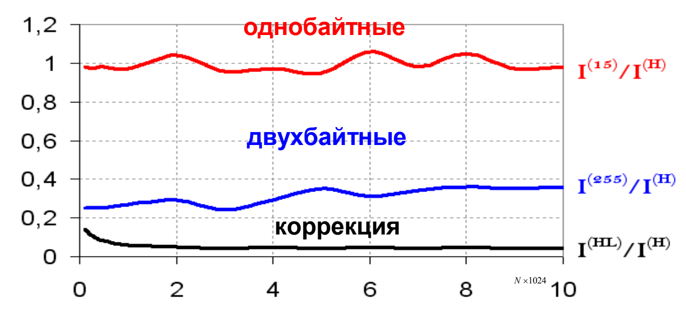


Рисунок 6 – Количество шагов на каждом этапе алгоритма в зависимости от размерности матрицы нейронных связей **A**

Видно, что на первом этапе $I^{(1)}/I^{(H)} \approx 1$. Это соотношение не зависит от выбора $q^{(1)}$, количество итераций на 2-м и 3-м этапах зависит от значения величины градаций на 1-м этапе: $I^{(2)}/I^{(H)} \approx 0{,}3$, а $I^{(HL)}/I^{(H)} \approx 0{,}05$. Количество итераций на 2-м этапе не зависит от величины градаций.

Быстродействие алгоритма зависит от выбора параметра $q^{(1)}$, применяемого на первом этапе. При увеличении числа градаций на 2-м этапе число шагов на этапе коррекции положения минимума исходной нейронной сетью сокращается до 1. Поэтому на втором этапе должно быть выбрано $q^{(2)} \geq 2048$ и 3-й этап коррекции может оказаться ненужным.

Таблица 1 – Ускорения алгоритма по сравнению со стандартным методом Хопфилда в зависимости от значений параметра $q^{(1)}$

Размерность сети $N = 256$	
Число градаций, q	Ускорение алгоритма, θ
2	3,3
4	4,3
8	7,9
12	10,3
15	11,2
32	6,4
64	3

В табл. 1 приведены результаты ускорения алгоритма по сравнению со стандартным методом Хопфилда в зависимости от значений параметра $q^{(1)}$. Результаты получены на матрице размерности $N = 256$ и $q^{(2)} = 255$ (наихудший вариант). Видно, что оптимум функции $\theta(q)$ более 10 раз достигается при $q^{(1)} = 10 \div 15$. Для расчета величины ускорения алгоритма по формуле (11) использовались данные, приведенные на рис. 6.

Заключение

Вероятность нахождения глобального минимума не зависит от числа градаций и та же, что и при использовании сети Хопфилда.

Модификация процедуры клиппирования матрицы нейронных связей позволяет ускорить вычислительный процесс более чем в 10 раз.

Найдены оптимальные значения числа градаций на каждом этапе работы алгоритма: $q^{(1)} = 12$, $q^{(2)} \geq 2048$. Результаты подтверждены для матриц размерности $N = 64 \div 10240$.

Литература

1. Van Hemmen J.L. Nonlinear neural networks near saturation / J.L. van Hemmen // Physical Review A. – 1987. – № 36. – P. 1959-1962.
2. Kintzel W. Models od Neural Networks I / W. Kintzel, M. Opper // Physics of Neural Networks / eds. E. Domany, J.L. van Hemmen, K. Schulten. – Springer, 1995. – P.170.
3. Widrow B. Adaptive switching circuits / B. Widrow, M.E.Jr. Hoff // IRE Western Electric Show and Convention Record. – 1960. – Part 4. – P. 96-104.
4. Алиева Д.И. Модель Хопфилда малых размеров с клиппированными связями / Д.И. Алиева, В.М. Крыжановский // Искусственный интеллект. – 2006. – № 3. – С. 240-248.
5. Крыжановский В.М. Клиппирование модели Хопфилда малых размеров / В.М. Крыжановский, Д.И. Симкина // Вестник компьютерных и информационных технологий. – 2007. – № 10. – С. 27-31
6. Крыжановский Б.В. О возможности применения процедуры клиппирования в задачах оптимизации / Б.В. Крыжановский, В.М. Крыжановский // IX Всероссийская научно-техническая конференция «Нейроинформатика-2007». – М. : МИФИ, 2007. – Т. 1. – С. 197-205.
7. Крыжановский Б.В. Применение процедуры клиппирования в задачах бинарной минимизации квадратичного функционала / Б.В. Крыжановский, В.М. Крыжановский, А.Л. Микаэлян // ДАН-2007. – 2007. – Т. 413, № 6. – С.730-733.

М.В. Крижановський, М.Ю. Мальсагов

Узагальнення процедури кліпування у задачах оптимізації у дискретному просторі

Досліджено можливість застосування процедури кліпування в задачі оптимізації квадратичного функціонала $E = (\mathbf{x}, \mathbf{Ax})$. Показано, що безпосереднє застосування процедури кліпування не дає особливого виграшу в прискоренні роботи алгоритму при пошуку глобального мінімуму. Запропоновано модифікацію процедури кліпування з параметром q (число градацій). Показано, що зі збільшенням q можливість співпадання напрямку градієнтів $E(x)$ та його кліпованого аналога $E_C(x) = (\mathbf{x}, \mathbf{Cx})$ зростає до 1.

M.V. Kryzhanovsky, M.U. Malsagov

Generalization of Clipping Procedure for Optimization Problems in Discrete Space

Capability of using clipping procedure for problem of optimization quadratic functional $E = (\mathbf{x}, \mathbf{Ax})$ was researched. It is shown application of clipping procedure doesn't give special benefit in acceleration of global minima search algorithm. Modification of clipping procedure with parameter q (the number of gradation) was suggested. It is shown probability of conjunction of gradients directions $E(x)$ and its clipped analogue $E_C(x) = (\mathbf{x}, \mathbf{Cx})$ raise to 1 with increasing of q .

Статья поступила в редакцию 20.05.2009.